Regionalizing Africa: Patterns of Precipitation Variability in Observations and Global Climate Models

Hamada S. Badr [*]

Department of Earth and Planetary Sciences (EPS)
Johns Hopkins University (JHU)
3400 N. Charles Street, Olin Hall, Baltimore, MD, 21218, USA
badr@jhu.edu

Amin K. Dezfuli

Climate and Radiation Laboratory (Code 613)
NASA Goddard Space Flight Center (NASA/GSFC)
8800 Greenbelt Road, Greenbelt, MD 20771, USA
and Universities Space Research Association
amin.dezfuli@nasa.gov

Benjamin F. Zaitchik

Department of Earth and Planetary Sciences (EPS)
Johns Hopkins University (JHU)
3400 N. Charles Street, Olin Hall, Baltimore, MD, 21218, USA
zaitchik@jhu.edu

Christa D. Peters-Lidard

Earth Sciences Division (Code 610)
NASA Goddard Space Flight Center (NASA/GSFC)
8800 Greenbelt Road, Greenbelt, MD 20771, USA
christa.d.peters-lidard@nasa.gov

[*] Corresponding author at: JHU, 3400 N. Charles Street, Olin Hall, Baltimore, MD, 21218, USA.
E-mail address: badr@jhu.edu (Hamada S. Badr).

1   ABSTRACT

2   Many studies have documented dramatic climatic and environmental changes that have

3   affected Africa over different timescales. These studies often raise questions regarding the spatial

4   extent and regional connectivity of changes inferred from observations, proxies, and/or derived

5   from climate models. Objective regionalization offers a tool for addressing these questions. To

6   demonstrate this potential, we present applications of hierarchical climate regionalizations of

7   Africa using observations and GCM historical simulations and future projections. First, we

8   regionalize Africa based on interannual precipitation variability using CHIRPS data for the period

9   1981-2014. A number of data processing techniques and clustering algorithms are tested to ensure

10 a robust definition of climate regions. These regionalization results highlight the seasonal and even

11 month-to-month specificity of regional climate associations across the continent, emphasizing the

12 need to consider time of year as well as research question when defining a coherent region for

13 climate analysis. CHIRPS regions are then compared to those of five GCMs for the historic period,

14 with a focus on boreal summer. Results show that some GCMs capture the climatic coherence of

15 the Sahel and associated teleconnections in a manner that is similar to observations, while other

16 models break the Sahel into uncorrelated subregions or produce a Sahel-like region of variability

17 that is spatially displaced from observations. Finally, we examine shifts in climate regions under

18 projected 21st century climate change for different GCMs and emissions pathways. We find a

19 projected change in the coherence of the Sahel, in which the western and eastern Sahel become

20 distinct regions with different teleconnections. This pattern is most pronounced in high emissions

21 scenarios.

1. Introduction

1.1 Climate Regionalization

Climate regionalization divides a region into homogeneous subregions based on one or more climatic variables. It is a very important step in climate studies because it helps in identifying the drivers of climate variability specific to each region (e.g., Dezfuli and Nicholson 2013; Nicholson and Dezfuli 2013). Applying conventional geographic boundaries for climate studies is often problematic because climate conditions and sensitivities can vary widely within a single study area (e.g., a country, a river basin). Standard climate classification systems (e.g., Köppen Climate Classification) also have limitations: (i) they represent mean climatic conditions rather than temporal variability (e.g., interannual), (ii) they are not informative for identifying drivers of variability, and (iii) as prescribed classification systems, rather than tools, they specify which variables and variable relationships are employed in the classification, rather than allowing the investigator to define characteristics of interest. Climate regionalization provides a useful alternative method for defining regions when we want to: (i) unravel drivers of climate variability specific to different regions and seasons (e.g., Badr et al. 2014a), and explore potential changes in the future, (ii) understand the spatial distribution of climate sensitivities, or (iii) employ a flexible system to understand how spatial patterns of variability differ for different climate variables (e.g. air temperatures as opposed to precipitation). These features make climate regionalization particularly valuable for applications that rely on identifying areas of common variability in a parameter of interest; for example, managing a climatically diverse hydrologic unit in the context of climate variability or change, filling in data gaps in the historic climate record, or optimizing seasonal forecast systems.

44   Climate regionalization applies an objective single- or multi-variate statistical technique

45   such as clustering (e.g., Burn 1989; Dezfuli 2011; Gong and Richman 1995; Isik and Singh 2008;

46   Ramachandra Rao and Srinivas 2006). Different clustering techniques have been applied in the

47   literature (Djomou et al. 2015; Herrmann and Mohr 2011; Janicot 1992; Mahe et al. 2001;

48   Nicholson 1986; Ogallo 1989), which are sensitive to conceptual approach, clustering algorithm,

49   data processing, and validation criteria. It is difficult to compare the regionalization results of

50   previous studies due to the lack of software tools that are technically designed for climate studies

51   and meet the preprocessing and postprocessing requirements. This has motivated us to develop an

52   open-source R package (Badr et al. 2014b) for Hierarchical Climate Regionalization (called,

53   "HiClimR"). Badr et al. (2015) describes the methodology and technical details of HiClimR.

54   The criteria used to interpret and validate climate regionalization may vary depending on

55   the study objectives. Here, our regionalization criteria are to find homogeneous regions that are

56   geographically contiguous, provided that the minimum region size is reasonable with respect to

57   the nature of the problem (i.e., it is consistent with general size constraints such as landscape

58   structure, data coverage and density, and known climate phenomena), and that the total number of

59   regions matches the inherent physical properties of interest (e.g., the number of regions for

60   identifying large-scale driver variability is different from dividing a country or area into regions

61   of coherent climate variability for national development). The optimum regionalization will

62   always have to involve some subjective decisions (e.g., contiguity checks and geographical

63   characteristics) since the problem is a combination of applications and statistics. However, the

64   objective criteria aim to maximize intra-regional correlations (i.e., the average correlations

65   between the region mean and all of its members) and minimize inter-regional correlations (i.e., the

66   correlations between region means).

1.2 Variability of African Precipitation

67

68   Africa is a continent of climate contrasts. The general spatial pattern of variability in mean

69   annual precipitation is widely familiar: a humid equatorial zone that includes the Congo forest,

70   transitional savannah zones as the tropics grade into the subtropics, subtropical deserts in the north

71   (Sahara) and southern (Kalahari) portions of the continent, and mid-latitude influences at northern

72   and southern extremities. Temporal precipitation variability in portions of Africa can be large and

73   is widely reported, in large part because of the social and economic impacts that hydroclimatic

74   extremes have had on the Sahel, the Horn of Africa, and several other regions.

75   From a climate dynamics perspective, the spatial and temporal variability of precipitation

76   present diverse challenges for process understanding, event prediction, and climate change

77   projection, and the climate vulnerability of many communities across Africa makes the problem

78   particularly urgent. In general, we understand variability to be a function of synoptic to mesoscale

79   atmospheric phenomena, including migration of the Intertropical Convergence Zone (ITCZ)

80   (Barry and Chorley 2009), the strength and location of atmospheric jets (e.g., the African Easterly

81   Jet and the Tropical Equatorial Jet) (Flohn 1964), monsoon circulations in West and East Africa,

82   and significant land-atmosphere interactions (Dickinson 1995), particularly in semi-arid zones.

83   These phenomena, in turn, are influenced by the sea surface temperatures (SST) in neighboring

84   oceans and by remote climate drivers that include the El Niño Southern Oscillation (ENSO), the

85   South Asian Monsoon, the Indian Ocean Dipole, the Atlantic Multidecadal Oscillation (AMO)

86   circulation, and many others. The relative influence of these dynamics and drivers varies in space

87   and time, and even the sign of influence of important drivers like ENSO can flip from season to

88   season or between subregions of a single country or river basin. This complexity can make it

89   difficult to characterize drivers of variability in a systematic way, or to track spatial changes in

90   their influence over time.

91   The purpose of this study is to identify regions within Africa that are coherent with respect

92   to interannual precipitation variability. This exercise offers an example of how regionalization can

93   be applied to characterize and study patterns of climate variability. It also provides a set of

94   regionalization results that can be applied to future studies of African climate. The regionalization

95   is performed using monthly precipitation estimates from observations and outputs from global

96   climate models (GCMs). Importantly, though perhaps not surprisingly, we find that regionalization

97   of Africa is a seasonally and even monthly specific problem. For this reason we define regions for

98   each season separately, where the season is defined as a combination of months for which regions

99   are spatially stable. We also find that regions differ when using different data sources (e.g.,

100   observations versus GCM). This paper presents the results of these seasonally and dataset-specific

101   regionalizations and explores implications for understanding drivers of interannual precipitation

102   variability and for projecting climate change using different GCMs.

103   2. Data

104   The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) dataset

105   version 2.0 (Funk et al. 2015) was used to represent the observed precipitation in the period 1981-

106   2014. The data are distributed on a 0.05° grid and a primary pentad temporal resolution with the

107   availability of aggregates (dekadal and monthly) or disaggregates (daily). In this study, we used

108   CHIRPS data for Africa at the monthly temporal resolution. The data were resampled to 0.25°

109   resolution due to the computational and memory requirements for regionalization of the entire

110   continent of Africa. In comparison with the observational datasets available, CHIRPS data

111   provides higher resolution, better station coverage over Africa, improved statistical approaches,

112   and updated temporal coverage.

113   Extended Reconstructed Sea Surface Temperature (ERSST) v4 (Huang et al. 2015; Liu et

114   al. 2015) monthly data ($2° \times 2°$ grid resolution; 1854-present) were used to test the correlation

115   patterns of each region's mean timeseries with global SST.

116   The outputs from five different GCMs (CCSM4, CNRM, GFDL, HadGEM2, and

117   MIROC5) were used to demonstrate the effectiveness of regionalization in the evaluation of GCMs

118   in terms of capturing the spatial patterns of precipitation variability and the evolution of regions in

119   response to greenhouse gas concentration as represented by coherent regions of the four

120   Representative Concentration Pathways (RCPs): RCPs, RCP2.6, RCP4.5, RCP6, and RCP8.5

121   (Moss et al. 2010).

3. Methods

123   Climate regionalization was performed to divide Africa into smaller regions that are

124   homogeneous with respect to interannual variability of precipitation for all months and 3-month

125   running average seasons. The Hierarchical Climate Regionalization (HiClimR) R package (Badr

126   et al. 2015) was used. The most appropriate clustering method was used for each case based on

127   three main criteria: homogeneity, separability, and contiguity. The homogeneity criterion

128   minimizes the within-region variability and is measured by the average correlation between the

129   region mean and its members (intra-regional correlation). The separability criterion maximizes the

130   differences between regions and is measured by the maximum correlation between the different

131   means (inter-regional correlation). The contiguity criterion visually identifies the member of each

132   region in geographical proximity not to have a region divided into distant subregions or members.

133   The method that provides higher overall homogeneity, lower inter-regional correlations, and

134   contiguous regions is referred as "better," even if the differences are small. In general, the different

135   methods will likely generate similar regions with slightly different statistics besides the visual

136   contiguity of the regions.

137   We choose the method with better statistics and better contiguity. Specifically, results from

138   two clustering methods are presented: Ward's method (Murtagh 1983; Ward Jr 1963), which

139   minimizes the error sum of squares between all members within a region after merging, and

140   Regional Linkage (Badr et al. 2015), which minimizes the inter-regional correlation between

141   region means at each merging step. Ward's method tends to generate well-proportioned regions

142   with high homogeneity regardless the inter-regional correlations, while regional linkage

143   emphasizes separation of systematically dissimilar regions. An added advantage of regional

144   linkage is that separating systematically dissimilar records isolates noise in the dataset—i.e.,

145   stations or very small clusters with completely different variability that cannot be merged into any

146   of the regions and are not correlated with other stations or clusters. The noisy stations or clusters

147   can be the result of bad data or may represent a phenomenon at a different scale (e.g., local effect).

148   For regionalization at continental scale, both forms of noise are undesirable as they are not

149   representative of the broad regions that are being defined. The isolation of that noise can also help

150   in the quality control of the data.

151   Several preprocessing options in HiClimR have been utilized to find the "optimal" regions

152   and to test the sensitivity of regionalization. Geographic masking was used to mask all stations

153   outside the continent of Africa. Grid cells with near-zero precipitation variability and/or very low

154   mean precipitation are masked out to avoid any negative impacts on the quality of regionalization.

155   The final regions are generated from detrended and standardized data to account only for the

156   interannual precipitation variability without any possible effects of the linear trend or precipitation

157   totals. The entire CHIRPS record available at the time of analysis (1981-2014) was used for all

158   regionalizations.

159   For GCM regionalization we focus on boreal Summer (July-September, JAS). We present

160   only one season in order to make the results digestible, and we choose JAS because of its

161   importance as the primary rainy season in the Sahel. Results are presented for the Historical

162   simulations of five GCMs (CCSM4, CNRM, GFDL, HadGEM2, and MIROC5; 1960-1990). A

163   unified time period of 30 years (1960-1990) was used for all GCMs; this period is used as a

164   baseline to compare the historical and future simulations of GCMs such as the reports of

165   Intergovernmental Panel on Climate Change (IPCC). The "Historical" simulations include historic

166   observations of greenhouse gases and other external forcings but are fully coupled to the ocean

167   and are not initialized from observations; as such climatological patterns and long-term trends are

168   expected to match historical patterns but specific year-to-year and even decade-to-decade

169   variability does not align with the observed climate record.

170   The possible effects of greenhouse gas concentration on spatial patterns of interannual

171   precipitation variability are examined by performing regionalization for the four Representative

172   Concentration Pathways (RCPs: RCP2.6, RCP4.5, RCP6, and RCP8.5) simulated by CCSM4

173   model for different 30-year periods within the 2006-2100 projection, and for the entire 2006-2100

174   time period. Additionally, to compare between the regions generated from observations (1981-

175   2014) and CCSM4 using the same time period, we combined the CCSM4 historical simulation

176   (which runs through 2005) with the RCP 4.5 projection (2006-forward) to create a 1981-2014

177   CCSM4 record. The RCP 4.5 emissions are reasonably consistent with observation for this period,

178   and the impact of emissions trajectory on climate response on such a short time scale is small

179   relative to internal variability. CHIRPS data were regridded to match the coarse model resolution.

4. Results & Discussion

4.1 Data Preprocessing

HiClimR implements several features to facilitate spatio-temporal analysis applications, including data filtering with geographic masking and/or mean/variance thresholds, data preprocessing via detrending and standardization. These were applied as described below.

4.1.1 Data Filtering

Fig. 1 shows the effect of masking options on regionalization results for the interannual variability of precipitation over Africa in January using Ward's clustering method, and Fig. 2 shows the associated clustering dendrograms for: no masking (Fig. 1A), geographic masking of Africa (Fig. 1B), and masking with filtering to remove all noncontiguous subregions, including Northern Africa (Fig. 1C). Filtering in this application included excluding the stations above 10N, in order to focus on sub-Saharan Africa, and removing small spatially discontinuous regions, such as Ethiopia in the result shown in Fig. 1C. This filtering is additional to the automatic filtering techniques in HiClimR that remove stations with near-zero variance and/or very low mean defined by a mean threshold (the mean threshold is typically selected as a very small fraction of the average monthly total precipitation where reasonable changes in its value do not affect the regionalization results; for CHIRPS data, a mean threshold of 12 mm/month was selected). The geographic masking improves the results because it excludes the artifacts introduced by Europe and small islands around Africa while filtering cleans up the regions and increases overall homogeneity.

Fig. 2 clarifies the regionalization quality in each case. The vertical axis represents the clustering height. The units of this axis depend on the clustering method, but the metric generally needs to be minimized to achieve maximum intra-regional homogeneity. For example, Region 4, which is mainly in Europe, has an artifact subregion in the border between Angola and Namibia

203   that disappears after geographic masking. For Ward's clustering, the y-axis on the dendrogram is

204   the sum of squared distances within all regions and is a measure of intra-regional variance. This is

205   different from the regional-linkage method that minimizes the maximum inter-regional correlation

206   between regions as a measure of region separation. To get the "optimal" number of regions for

207   final regionalization, we need to minimize the inter-regional correlations (region separation) and

208   maximize the intra-regional correlation (region homogeneity). Hence, we add a horizontal axis on

209   the left of the dendrogram plots in Fig. 2 to show the maximum inter-regional correlation at each

210   dendrogram cut that defines the number of regions. It is clear that the overall clustering height

211   decreases when applying the geographic masking from Fig. 2A to Fig. 2B and further with filtering

212   the data from Fig. 2B to Fig. 2C, which is consistent with the regionalization qualities in Fig. 1A-

213   C.

214   We emphasize that our selection of geographic extent and masking procedure is specific to

215   the objectives of this study—to provide a stable and informative regionalization of Africa at

216   continental scale. Regionalization could just as easily be applied to a global domain, including

217   land and ocean, in order to study global scale response to major modes of climate variability, for

218   example. This would change the regionalization of Africa but might yield other insights on climate

219   variability.

220   *4.1.2 Detrending and Standardization*

221   Fig. 3 shows the effects of data detrending and standardization on regionalization results

222   for the interannual variability of precipitation over Africa in January using Ward's clustering

223   method. Fig. 4 shows the associated clustering dendrograms. We found that standardization had

224   no visible effect on regionalization results in this month (not shown; this may be different when

225   using a different data set or at another temporal scale, when there is large spatial variability in

226 precipitation magnitudes), while detrending affects regions that have strong similarities or

227 differences in linear trends. In this application, detrending tends to shift the borders between

228 regions without fundamentally altering the character of the map (Fig. 3). But these shifts are

229 systematic: for example, regions 3 and 4 share a positive linear trend that tends to increase the

230 correlation between regions if detrending is not applied. These regions also show the highest inter-

231 regional correlation in the analysis, so the maximum inter-regional correlation figures shown on

232 the left side of Fig. 4 reflect higher, trend-influenced correlations between these regions in the raw

233 data (0.58) and a lower correlation when data are detrended (0.48). Note that the region merging

234 in Fig. 4 is based on Ward's method, which minimizes the variance within regions.

235 **4.2 Dissimilarity Measures**

236   The dissimilarity measure—or the nature of the temporal dimension used to regionalize

237 spatial data in HiClimR—is a crucial decision, and the choice depends on the specific application.

238 For example, regions can be generated based on interannual, intraseasonal, or daily variability or

239 on seasonal cycle. Fig. 5 shows 12 regions generated for precipitation over Africa using Ward's

240 clustering method based on the interannual variability (Annual Mean; left) and seasonality (Annual

241 Cycle; right). It is clear that the differences in seasonality don't always align with differences in

242 interannual variability: regions based on seasonality tend to align zonally, following the seasonal

243 migration of the intertropical convergence zone within tropical Africa, while those based on the

244 annual mean show both zonal and meridional structure, reflecting differing influences of remote

245 climate drivers.

246 **4.3 Monthly-Specific Regions**

247   Fig. 6 shows results of regionalization on interannual precipitation variability performed

248 separately for each month of the year. This is an optimized version of regionalization that applies

11

249 masking, filtering, standardization and detrending, as described above. These options were

250 selected for this application based on the sensitivity analysis described in Section 4.1. Ward's

251 method provided the best regionalization results in winter months (December-March) and very

252 similar regions in May. However, it was sensitive to region size, which sometimes results in

253 dividing a large but highly homogeneous region into two or more regions. In contrast, the regional

254 linkage method is able to identify big homogenous regions like the Sahel in summer and to filter

255 out noisy data. In winter months/seasons, the regions have relatively similar size and Ward's

256 method performs well: even though the method optimizes for intra-regional homogeneity rather

257 than inter-regional separability, the correlation between regions was reasonably low, indicating the

258 separation of regions was meaningful. In summer and most transitional months, however, regional

259 linkage provides more climatically meaningful regions, while Ward's tends to split regions based

260 on size, producing multiple regions that have high inter-regional correlation. In May, both methods

261 generated very similar regions (not shown) and Ward's method was selected for the slightly better

262 overall homogeneity. Note that the numbering of regions is arbitrary; the regionalization process

263 simply distinguishes between regions based on dissimilarity measure, and the analyst must

264 interpret association of regions across datasets or regionalization methods.

265   It is clear that the spatial patterns of precipitation variability over Africa (as represented by

266 homogeneous regions in Fig. 6) are specific to calendar month. This can be used as a guideline for

267 researchers studying climate processes, as one would not want to average across months with

268 significantly different regionalization patterns in the area of interest, just as one would not want to

269 average across two poorly correlated regions in any given season. For example, for a large-scale

270 study these results suggest that January and February can be treated as a coherent season, with

271 only small changes in region boundaries in central Africa. Moving back to December or forward

272   to March, however, regions in southern Africa begin to split and shift relative to January-February

273   in ways that merit attention before attempting to generalize across DJF or JFM in those areas. This

274   could have implications for forecasting.

275   Table 1 lists the average intra-regional correlations (between region mean and all members)

276   for all regions and the maximum inter-regional between region means. The winter months

277   (December-March) and May utilized Ward's method and their rows are highlighted in bold font.

278   All other months, especially the summer months (July-September) and months of complicated

279   variability in transition between seasons utilize the regional linkage method and its ability to isolate

280   noisy areas for removal. In other applications, where optimal quality control of the data is desired,

281   statistically isolated regions (or weather stations) could be treated in a statistical or dynamical

282   analysis to understand reasons for isolation and fill their gaps. The values of intra-regional

283   correlations in Table 1 are affected by the very high resolution and continental scale (using coarser

284   data would increase the overall homogeneity of the regions since larger grid cells are smoothed

285   averages of the included finer grid cells), which indicates that each of the regions can be divided

286   into smaller regions for finer scale applications such as hydrological analysis over one country or

287   a smaller region of interest. However, the current results target the association of sub-continental

288   regions with large-scale drivers of variability (teleconnections). The maximum inter-regional

289   correlation indicates the correlation between the *most similar* regions in the regionalizations. All

290   other regions have smaller or negative correlations between means.

291   **4.4 Historical vs Future (GCMs)**

292   Fig. 7 shows the regions of Africa based on interannual variability of summer (July-

293   September, JAS) precipitation using observations from CHIRPS (OBS; 1981-2014) and different

294   GCM historical simulations (CCM4, CNRM, GFDL, HadGEM2, and MIROC5; 1960-1990). A

13

295 unified time period of 30 years (1960-1990) was used for all GCMs. To avoid regionalization

296 sensitivity based on clustering approach, all models are treated similarly regarding the number of

297 regions (cut-off level of the dendrogram), data thresholds (variance and mean thresholds for data

298 filtering), and preprocessing options (e.g., detrending and standardization). The regional linkage

299 method is appropriate for summer precipitation when using CHIRPS observations, the outputs

300 from the two models that exhibit similar spatial variability to observations (CCSM4 and

301 MIROC5), and CNRM. In contrast, Ward's method was relatively better for GFDL and HadGEM2

302 as it provides slightly better overall homogeneity and contiguity, and none of the clustering

303 methods or preprocessing options had a significant impact on the dominant patterns. The observed

304 spatial patterns show the Sahel as one homogeneous region with strong agreement between the

305 regions generated from observations: the Sahel region is homogeneous (intra-regional correlation

306 = ~0.65), and independent from the other three regions. It is found that CCSM4 and MIROC5

307 have good skill in capturing the coherence of the Sahel precipitation signal in summer (strong

308 signal of unique variability over the big green region that represents the Sahel, which cannot be

309 divided into smaller subregions or the mean timeseries of subregions will be strongly correlated),

310 while the other models miss this coherence and divide the Sahel into smaller regions with

311 dissimilar interannual variability. CNRM generates random spatial patterns, GFDL simulates

312 precipitation in summer shifted in the northwestern direction and HadGEM2 divides the Sahel

313 region into eastern and western subregions. The observational analysis was repeated using CRU

314 in place of CHIRPS; results were similar and are not shown.

315   The robustness of Sahel region in Summer (JAS) was examined using regionalization of

316 different observational data sources (CHIRPS, CRU, and GPCC) and a variety of homogeneity

317 checks such as correlation patterns between the region mean and precipitation over Africa. All

14

318   data sources yield similar spatial patterns and identify the Sahel as a single coherent region.

319   Increasing the number of clusters or changing the clustering algorithm divides Sahel into two or

320   more "very similar" regions (i.e., the inter-regional correlations between the subregions are high

321   meaning that we need to merge them into one region).

322   Fig. 8 shows the regionalizations of Africa based on interannual variability of JAS

323   precipitation in 1960-1990 for six different ensemble members of CCSM4. All ensemble members

324   identify the Sahel and East Africa as homogeneous regions, consistent with the observations. The

325   known West African dipole mode is also detected in all cases. The differences primarily appear in

326   West Equatorial Africa, which has a strong intrinsic heterogeneity with respect to interannual

327   variability (Balas et al. 2007; Dezfuli and Nicholson 2013).

328   To compare between the regions generated from observations (1981-2014) and CCSM4

329   using the same time period, we combine the CCSM4 historical simulation (which runs through

330   2005) with the RCP 4.5 projection (2006-forward) to create a 1981-2014 CCSM4 record. The RCP

331   4.5 emissions are reasonably consistent with observation for this period, and the impact of

332   emissions trajectory on climate response on such a short time scale is small relative to internal

333   variability. CHIRPS data were regridded to match the coarse model resolution. It is found the

334   spatial patterns of interannual precipitation variability captured by CCSM4 in 1981-2014 (Fig. 9),

335   especially over the Sahel, are very similar to the unified period (1960-1990) in Fig. 7. In the next

336   sections, we focus on CCSM4 for the GCM and use the unified period to facilitate comparison

337   with CHIRPS.

338   Fig. 10 shows the correlation patterns of the four regions generated based on the interannual

339   variability of JAS precipitation using CHIRPS observations with global ERSST for the mean

340   timeseries of the region and Fig. 11 shows the correlations patterns from CCSM4 with its own

15

341   SST (both 1981-2014). Similarities and differences between observed regions and CCSM4 regions

342   are strongly related to SST teleconnections. Region 3 of CCSM4, which is the largest difference

343   between the observational and GCM-based regionalization, shows a positive correlation with the

344   SSTs over the ENSO region. However, such an association does not exist for any of the

345   observation-based regions; it appears to be a false ENSO teleconnection that exists only in the

346   model. For region 1, which corresponds to the Sahel in both model and simulations, the correlation

347   with North Atlantic SST (i.e., Atlantic Multidecadal Oscillation or AMO) is markedly different

348   and even slightly reversed for CCSM4 (Fig. 11) relative to observations (Fig. 10). In addition, the

349   model shows significant Region 1 correlations with Gulf of Guinea SSTs that are not seen in

350   observations. Together these differences suggest that Sahel sensitivity to Atlantic Ocean variability

351   is substantially different in CCSM4 than it is in observation. Region 2 of CCSM4 in Fig. 11, along

352   the Guinean coast, shows weaker correlation with SSTs of adjacent waters than the observations.

353   These agreements or disagreements between the regions of the observed and simulated

354   precipitation raise many interesting questions about the underlying mechanisms and skills of the

355   model dynamics and need further investigation.

356   The regions do not respond to unique climate drivers, but they have different variability in

357   response to a set of large-scale and local drivers. Regionalization helps in identifying regions with

358   coherent variability that are different from each other's. The climate drivers of a specific region

359   can then be identified, which are greatly improved over the commonly used geographic boxes that

360   could mix inhomogeneous regions and create unrealistic variability and associated drivers (e.g.,

361   the mean of an area defined by a box without performing regionalization to test its homogeneity

362   may include different –or perhaps opposite– variability that can ruin the analysis).

363   **4.5 Climate Projections**

364   Fig. 12 shows the possible effect of greenhouse gas concentrations on spatial patterns of

365   interannual precipitation variability as represented by coherent regions simulated by CCSM4

366   model for the period 2006-2100 for each of the four Representative Concentration Pathways

367   (RCPs: RCP2.6, RCP4.5, RCP6, and RCP8.5). In these simulations the Sahel falls into two

368   regions: regions 1 and 2 in Fig. 12. As radiative forcing increases (RCP2.6 through RCP8.5),

369   region 2 shrinks and region 1 grows: essentially, there is a shift from a coherent Sahel band that

370   stretches almost across the continent (region 2 in RCP2.6) to two distinct eastern and western Sahel

371   regions in higher emissions scenarios. The changes in Equatorial West Africa and East Africa

372   (regions 3 and 4) are small. We note that these 2006-2100 regionalization results average across

373   variability in regions that is observed in shorter periods of analysis. The evolution of regions for

374   the four climate projections examined at three 30-years different simulation periods (2010-2040,

375   2040-2070, and 2070-2100) suggest that the spatial patterns vary throughout the 21st century,

376   perhaps as a response to changes in global SST-rainfall relationships, but that there is a gradual

377   trend towards a split between the eastern and western Sahel (not shown).

378   Fig. 13 and Fig. 14, respectively, show the correlation patterns of CCSM4 JAS region 1

379   (western Sahel) and region 2 (eastern Sahel) precipitation from all RCPs with the corresponding

380   SST field for the entire simulation period (2006-2100). Teleconnections in low emissions scenarios

381   (RCP2.6, RCP4.5) are, as expected, more similar to Sahel teleconnections in historical simulations.

382   This is particularly evident for the eastern Sahel (region 2; Fig. 14), which approximately

383   corresponds to the unified Sahel in historical simulations (Fig. 9). For all emissions scenarios,

384   region 1 shows weaker correlation with global SST patterns than region 2 does, but the strength of

385   these connections increases as the region expands under higher emissions. Region 2 shows strong

386 correlations with the ENSO region and with the Indian Ocean under all RCPs, though the

387 relationship is strongest for low emissions.

388   Interestingly, the influence of tropical South Atlantic and Gulf of Guinea SSTs on both

389 regions changes as a function of emissions, shifting from a positive correlation at low emissions

390 to a neutral or negative correlation at high emissions. This suggests changing dynamical

391 interactions between remote forcings like ENSO and the more local influence of SSTs in the

392 neighboring Atlantic Ocean. It is also noteworthy that for region 2 there is an Atlantic Meridional

393 Mode (AMM) type signal in the historical simulations and RCPs 2.6 and 4.5, but this signal

394 disappears in RCP 8.5, where the cross-equatorial SST gradient is absent. This again points to

395 changing relationships between the tropical Atlantic Ocean and Sahel precipitation. We do note

396 that all results shown here are for a single GCM ensemble member, which is expected to impact

397 results for shorter time scales and might also impact the details of these 2006-2100 results.

5. Conclusions

399   The different clustering methods available in HiClimR can be useful for different

400 regionalization problems. Ward's method is sensitive to region size and tends to divide a large

401 homogeneous region into multiple separate regions. In contrast, the regional linkage method can

402 identify a big coherent region and filter out noisy data. For example, Ward's method was effective

403 in winter months, when regions have relatively similar size, while regional linkage yielded cleanly

404 separated and contiguous homogeneous regions in summer, when the interannual variability of

405 Sahel precipitation is dominant. Regional linkage also provides more climatically relevant results

406 in the transition seasons.

407   In the historical observational record, the Sahel expands from West to East and dominates

408 the interannual variability of African precipitation in summer. This is confirmed with the spatial

409 correlation patterns of the region mean and precipitation over Africa using CHIRPS data. The

410 overall homogeneity of the region is only moderately high, but it is spread across the region with

411 highly significant correlation between any potentially identifiable subregions. This suggests that

412 interannual variability is characterized by one dominant mode of variability for the entire region

413 plus local modes of variability in the subregions. The winter months (December-March) are

414 relatively stable in their regionalization, and with a few exceptions can be treated as a coherent

415 season for climate analysis. As expected, the transitional months have complicated spatial

416 variability.

417   We tested the potential of climate regionalization for model intercomparision and

418 assessment. CCSM4 and MIROC5 showed good skill in capturing the precipitation signal over the

419 Sahel in summer, while the other models miss the spatial patterns by dividing the Sahel into smaller

420 regions with dissimilar interannual variability. CNRM generates random spatial patterns, GFDL

421 simulates Sahel-like precipitation variability in summer shifted to the northwest and HadGEM2

422 divides the Sahel region into eastern and western subregions. This does not mean that models like

423 CNRM, GFDL, or HadGEM2 are not useful for projecting climate change in Africa. On the

424 contrary, the regionalization analysis presented in this paper might allow us to interpret these

425 simulations properly for studies of climate process and projections of climate change impacts. If

426 GFDL systematically shifts the "Sahel" climate pattern to the northwest, then analysis of Sahel

427 sensitivity in this model should be performed with a corresponding shift. This opens the possibility

428 of a more meaningful, regionalization-based multi-model ensemble of climate projections for the

429 Sahel—or for any region in which models differ in the spatial representation of variability. In the

430 long term, these studies can also inform model development to correct spatial biases and

431 displacements in the representation of climate variability.

432 *Acknowledgements*

References

441

442 Badr, H. S., B. F. Zaitchik, and S. D. Guikema, 2014a: Application of Statistical Models to the

443 Prediction of Seasonal Rainfall Anomalies over the Sahel. *Journal of Applied Meteorology and*

444 *Climatology*, **53,** 614-636.

445 Badr, H. S., B. F. Zaitchik, and A. K. Dezfuli, 2014b: HiClimR: Hierarchical Climate

446 Regionalization. *Comprehensive R Archive Network (CRAN)***,** [Available online at http://cran.r-

447 project.org/package=HiClimR].

448 ——, 2015: A tool for hierarchical climate regionalization. *Earth Science Informatics***,** 1-10.

449 Balas, N., S. Nicholson, and D. Klotter, 2007: The relationship of rainfall variability in West

450 Central Africa to sea-surface temperature fluctuations. *International journal of climatology*, **27,**

451 1335-1349.

452 Barry, R. G., and R. J. Chorley, 2009: *Atmosphere, weather and climate.* Routledge.

453 Burn, D. H., 1989: Cluster analysis as applied to regional flood frequency. *Journal of Water*

454 *Resources Planning and Management*, **115,** 567-582.

455 Dezfuli, A. K., 2011: Spatio-temporal variability of seasonal rainfall in western equatorial

456 Africa. *Theoretical and applied climatology*, **104,** 57-69.

457 Dezfuli, A. K., and S. E. Nicholson, 2013: The Relationship of Rainfall Variability in Western

458 Equatorial Africa to the Tropical Oceans and Atmospheric Circulation. Part II: The Boreal

459 Autumn. *Journal of Climate*, **26**.

460 Dickinson, R. E., 1995: Land-atmosphere interaction. *Reviews of Geophysics*, **33,** 917-922.

461 Djomou, Z. Y., D. Monkam, and R. Chamani, 2015: Characterization of climatic zones,

462 variability and trend in northern Africa. *Climate Dynamics*, **44,** 3481-3491.

463 Flohn, H., 1964: *Investigations on the tropical easterly jet.* Dümmlers Vlg.

464   Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—a new

465   environmental record for monitoring extremes. *Scientific data*, **2**.

466   Gong, X., and M. B. Richman, 1995: On the application of cluster analysis to growing season

467   precipitation data in North America east of the Rockies. *Journal of Climate*, **8,** 897-931.

468   Herrmann, S. M., and K. I. Mohr, 2011: A continental-scale classification of rainfall seasonality

469   regimes in Africa based on gridded precipitation and land surface temperature products. *Journal*

470   *of Applied Meteorology and Climatology*, **50,** 2504-2513.

471   Huang, B., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature Version 4

472   (ERSST. v4). Part I: Upgrades and Intercomparisons. *Journal of Climate*, **28,** 911-930.

473   Isik, S., and V. P. Singh, 2008: Hydrologic regionalization of watersheds in Turkey. *Journal Of*

474   *Hydrologic Engineering*, **13,** 824-834.

475   Janicot, S., 1992: Spatiotemporal variability of West African rainfall. Part I: Regionalizations

476   and typings. *Journal of Climate*, **5,** 489-497.

477   Liu, W., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature Version 4

478   (ERSST. v4): Part II. Parametric and Structural Uncertainty Estimations. *Journal of Climate*, **28,**

479   931-951.

480   Mahe, G., Y. L'hote, J. C. Olivry, and G. Wotling, 2001: Trends and discontinuities in regional

481   rainfall of West and Central Africa: 1951–1989. *Hydrological Sciences Journal*, **46,** 211-226.

482   Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research

483   and assessment. *Nature*, **463,** 747-756.

484   Murtagh, F., 1983: A survey of recent advances in hierarchical clustering algorithms. *The*

485   *Computer Journal*, **26,** 354-359.

486   Nicholson, S. E., 1986: The spatial coherence of African rainfall anomalies: interhemispheric

487   teleconnections. *Journal of climate and applied meteorology*, **25,** 1365-1381.

488   Nicholson, S. E., and A. K. Dezfuli, 2013: The Relationship of Rainfall Variability in Western

489   Equatorial Africa to the Tropical Oceans and Atmospheric Circulation. Part I: The Boreal

490   Spring. *Journal of Climate*, **26**.

491   Ogallo, L., 1989: The spatial and temporal patterns of the East African seasonal rainfall derived

492   from principal component analysis. *International Journal of Climatology*, **9,** 145-167.

493   Ramachandra Rao, A., and V. Srinivas, 2006: Regionalization of watersheds by hybrid-cluster

494   analysis. *Journal of Hydrology*, **318,** 37-56.

495   Ward Jr, J. H., 1963: Hierarchical grouping to optimize an objective function. *Journal of the*

496   *American statistical association*, **58,** 236-244.

497 # **Tables**

499   **Table 1** Intra-regional and maximum inter-regional correlations for each month.

MM	Intra-regional Correlations (R_{nn}[1])												R_{mx}[2]
	R_{01}	R_{02}	R_{03}	R_{04}	R_{05}	R_{06}	R_{07}	R_{08}	R_{09}	R_{10}	R_{11}	R_{12}	
01	**0.71**	**0.49**	**0.66**	**0.61**	**0.61**								**0.48**
02	**0.72**	**0.51**	**0.68**	**0.64**	**0.53**								**0.43**
03	**0.59**	**0.49**	**0.74**	**0.60**	**0.63**	**0.62**							**0.42**
04	0.57	0.72	0.66	0.67	0.73	0.87	0.70	0.74	0.57	0.72	0.67	0.73	0.49
05	**0.54**	**0.67**	**0.46**	**0.64**	**0.63**	**0.57**	**0.69**						**0.49**
06	0.54	0.46	0.50	0.57									0.35
07	0.60	0.58	0.54	0.66									0.39
08	0.58	0.61	0.62	0.60	0.59								0.42
09	0.52	0.56	0.54	0.59	0.61								0.36
10	0.69	0.67	0.64	0.68	0.56	0.82	0.72	0.64	0.73	0.70	0.80		0.51
11	0.54	0.68	0.70	0.60	0.63	0.67	0.68	0.70					0.44
12	**0.58**	**0.47**	**0.73**	**0.57**	**0.67**	**0.68**							**0.33**

500

---

[1] *R_{nn} is the average intra-regional correlation for region number nn, between the region mean and all members within the region.*
[2] *R_{mx} is the maximum inter-regional correlation between region means.*

Figures

534

Fig. 1 The effect of geographic masking and data filtering on the quality of regionalization: A)

no masking, B) geographic masking of Africa, and C) geographic masking and data filtering.

537
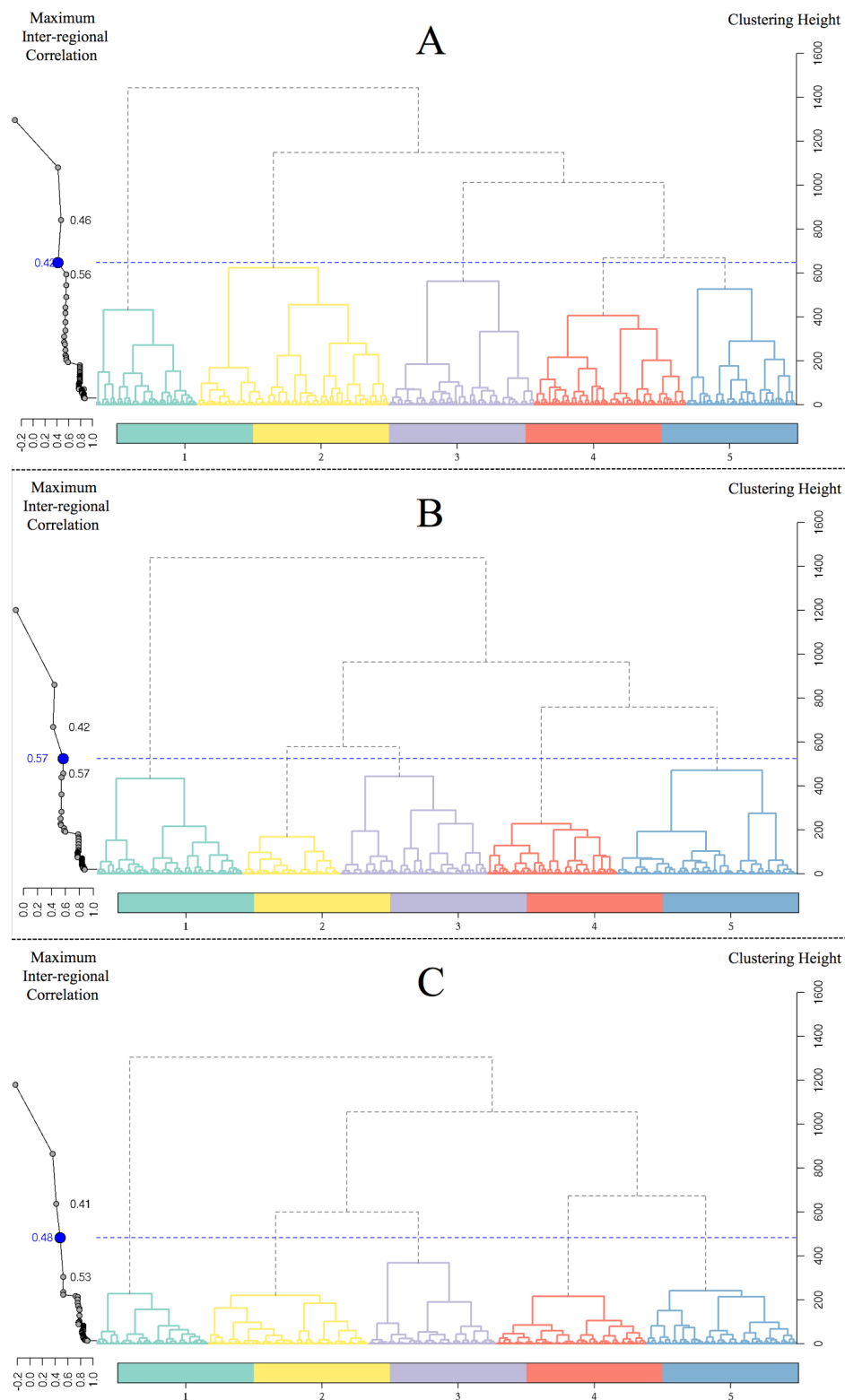
Fig. 2 The effect of geographic masking and data filtering on the clustering dendrogram: A) no

masking, B) geographic masking of Africa, and C) geographic masking and data filtering.
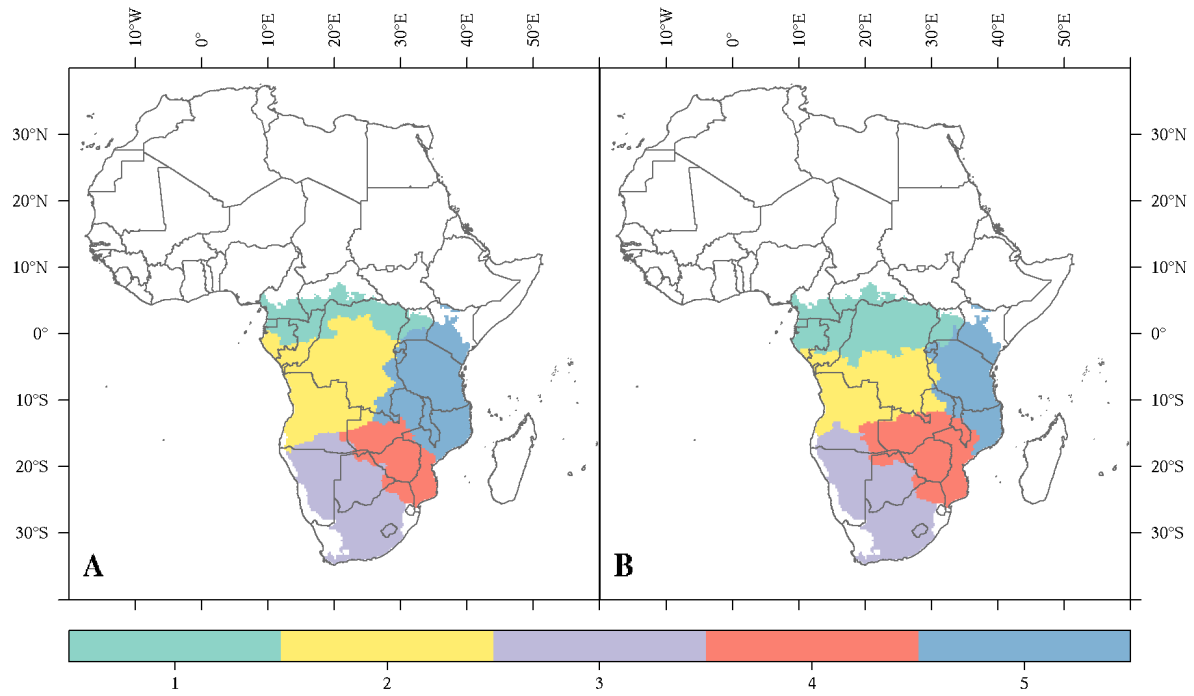
Fig. 3 The effect of detrending and standardization on regionalization quality: A) raw data and B) detrended and standardized data.
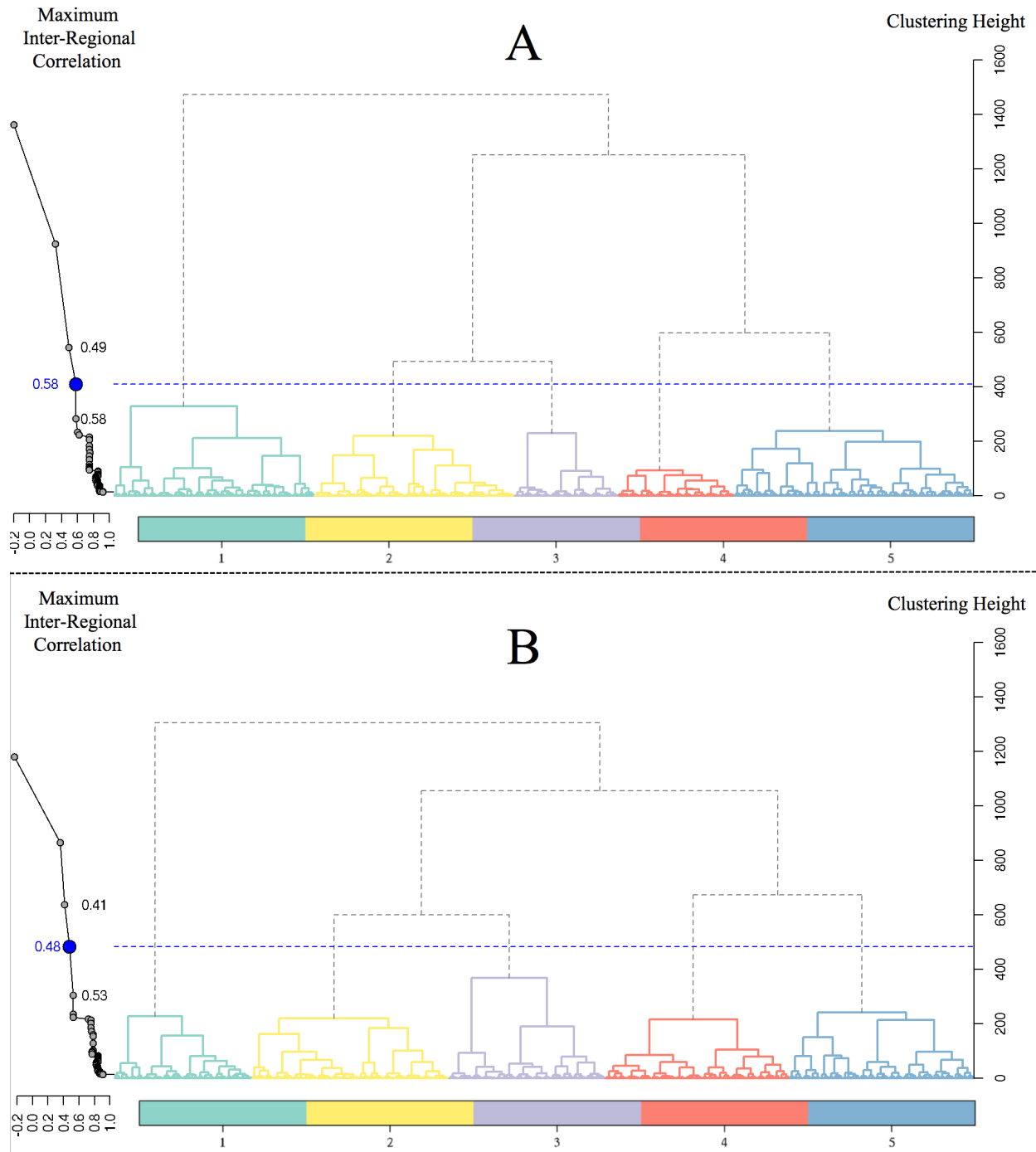
Fig. 4 The effect of detrending and standardization on the clustering dendrogram: A) raw data and B) detrended and standardized data.

546
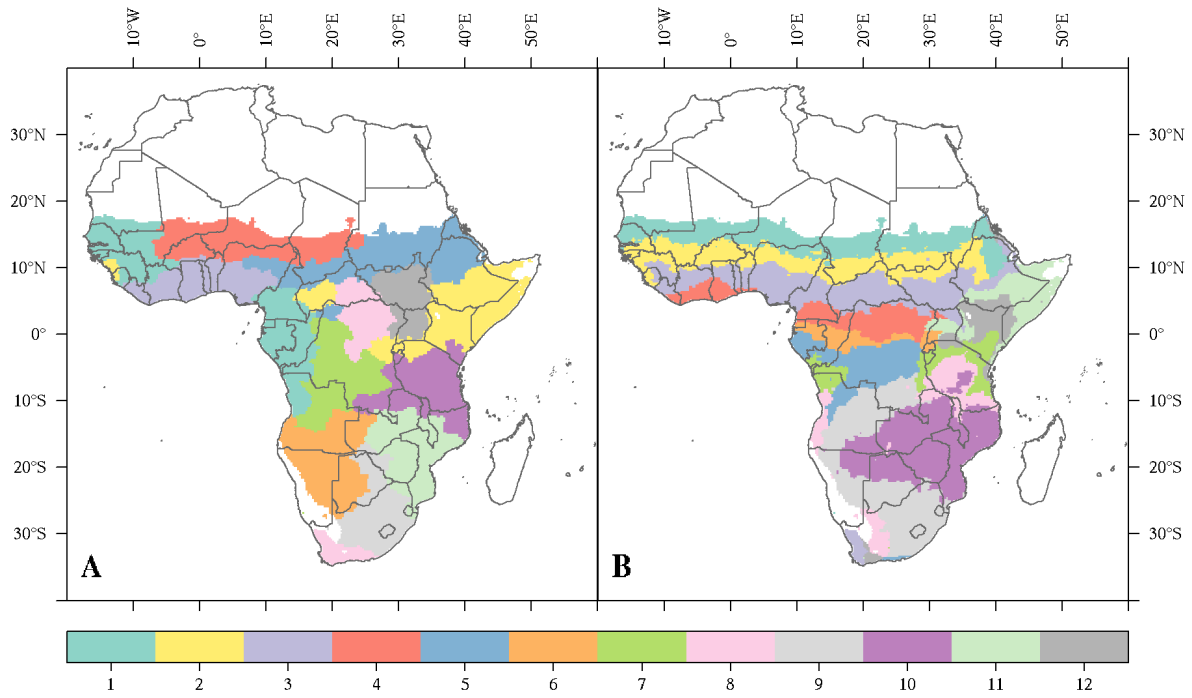
547 **Fig. 5** Regionalization based on: A) interannual variability of annual totals of precipitation and

548 B) annual cycle of precipitation over Africa using CHIRPS data v2.0 (1981-2014).
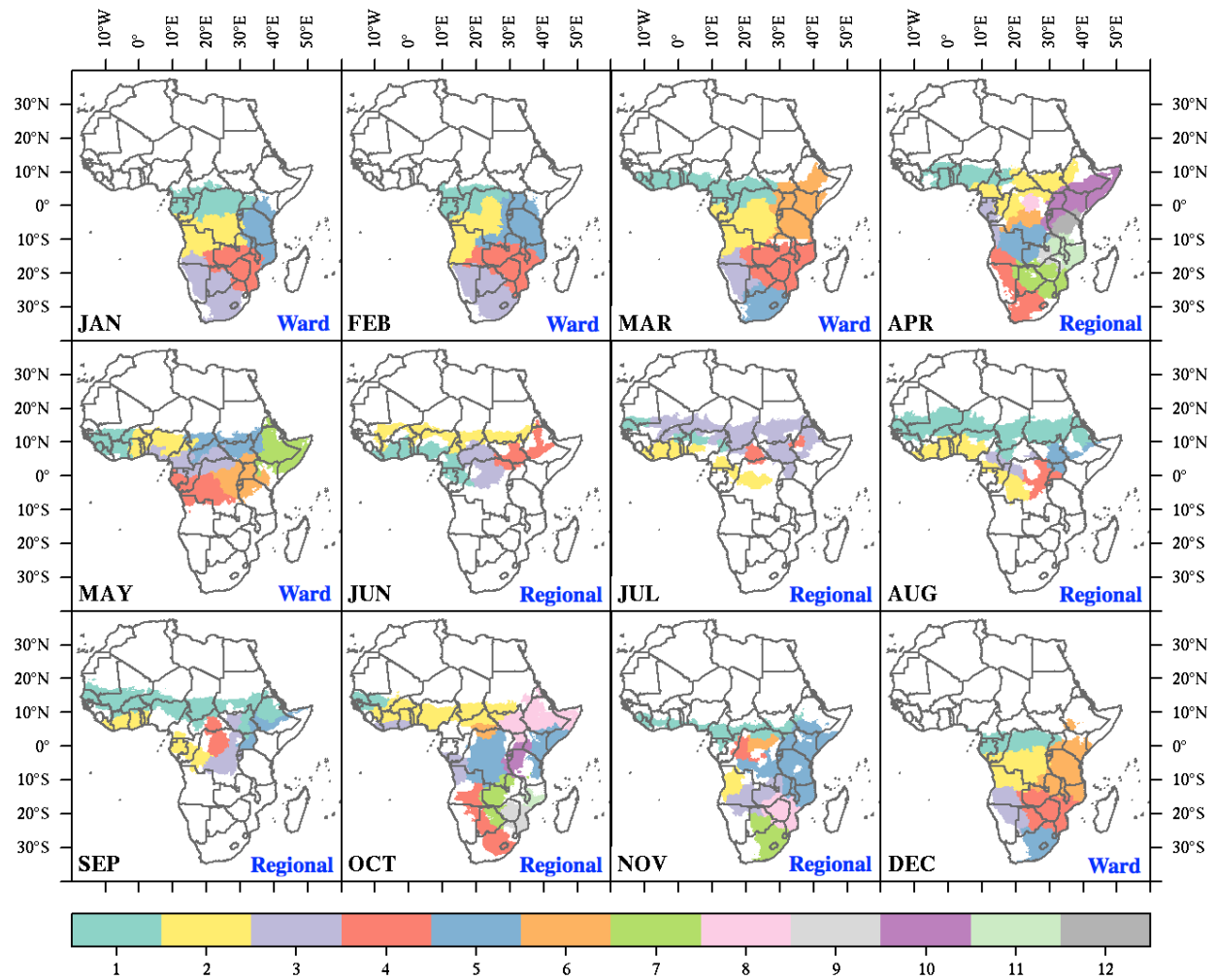
Fig. 6 Regionalization of Africa based on interannual variability of monthly precipitation.

549

550
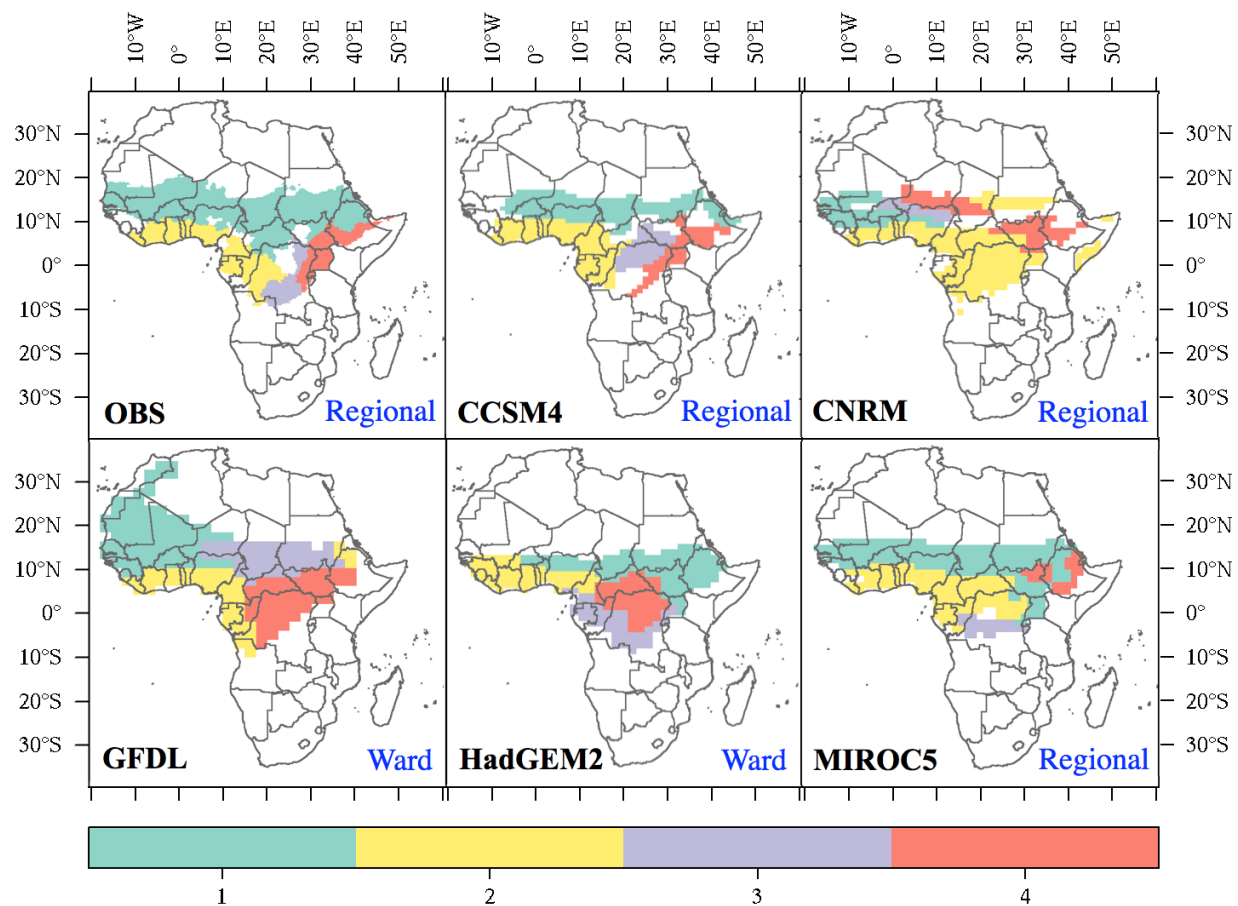
Fig. 7 Regionalization of Africa based on interannual variability of Summer (JAS) precipitation using CHIRPS observations (1981-2014) and different GCMs (1960-1990).

554
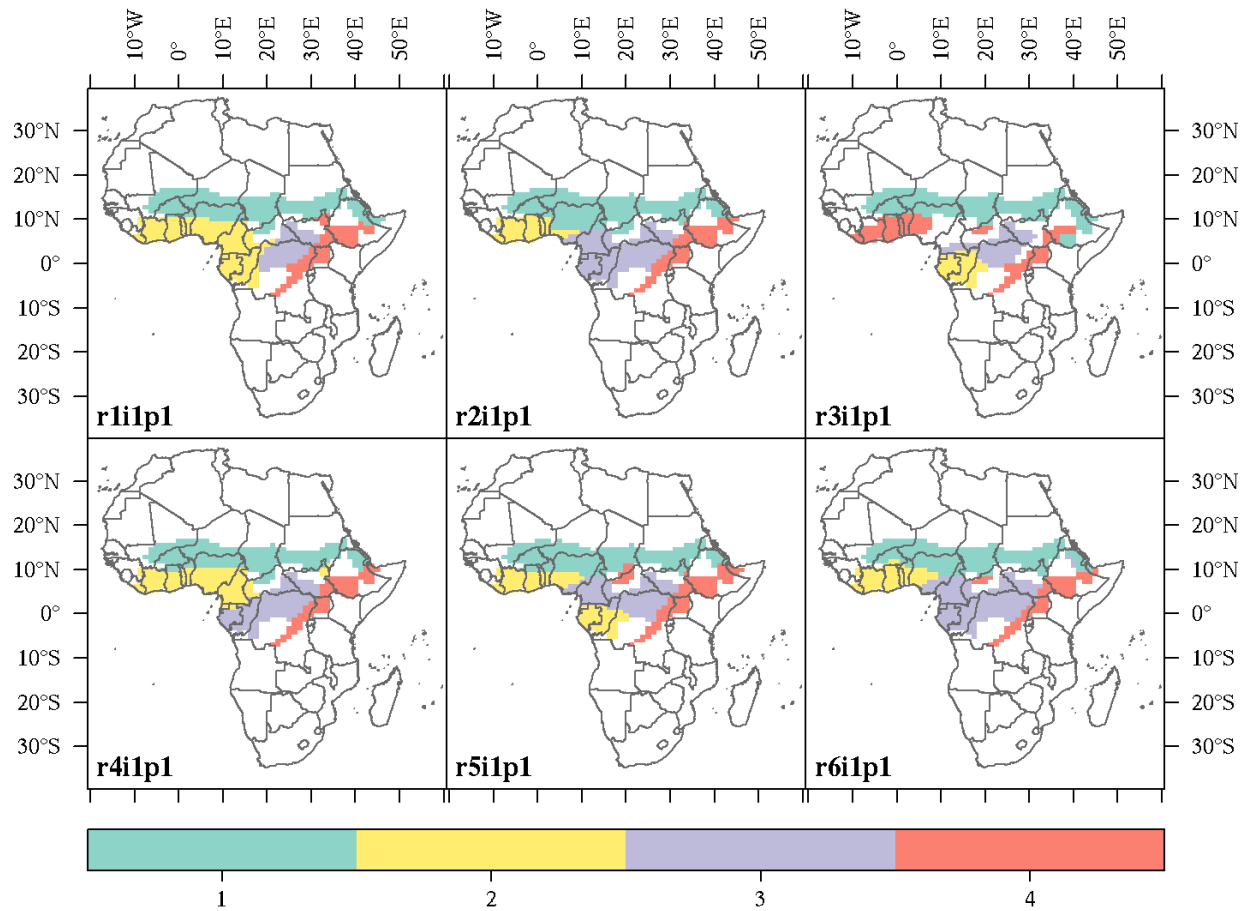
Fig. 8 Regionalization of Africa based on interannual variability of JAS precipitation in 1960-

1990 for different ensemble members of CCSM4.

557
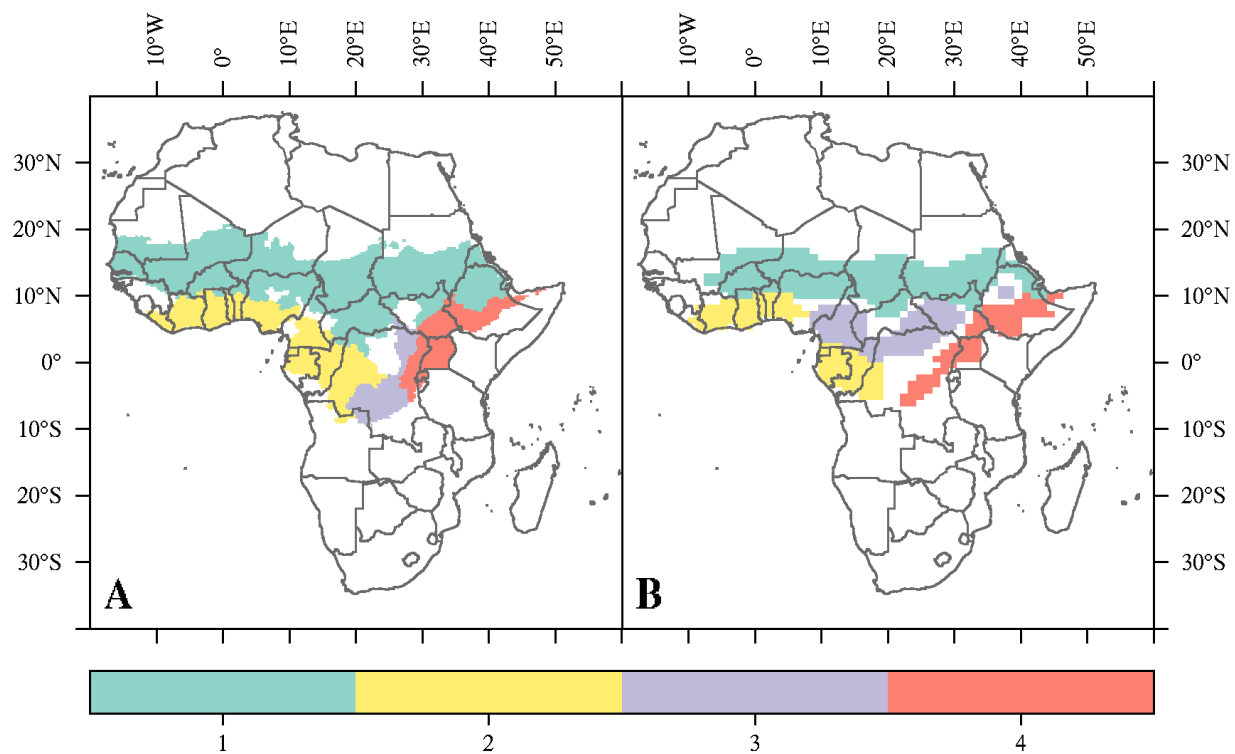
Fig. 9 Regionalization of Africa based on interannual variability of JAS precipitation in 1981-2014 using regional-linkage method: A) CHIRPS and B) CCSM4.
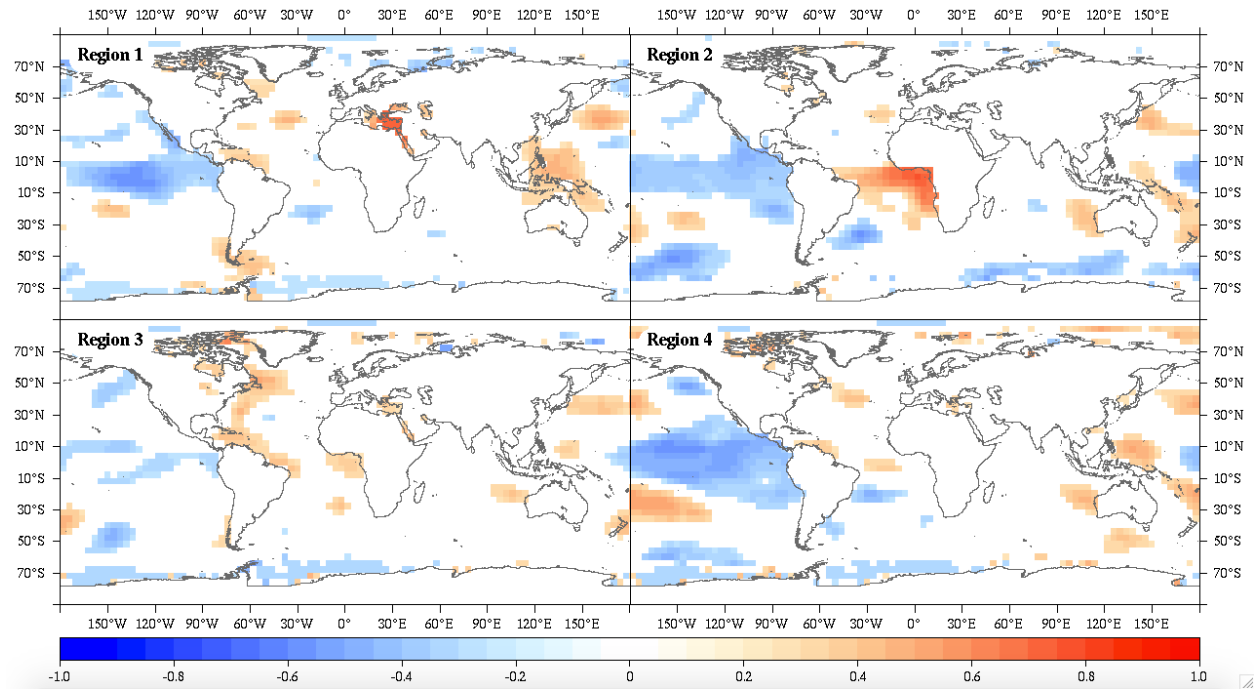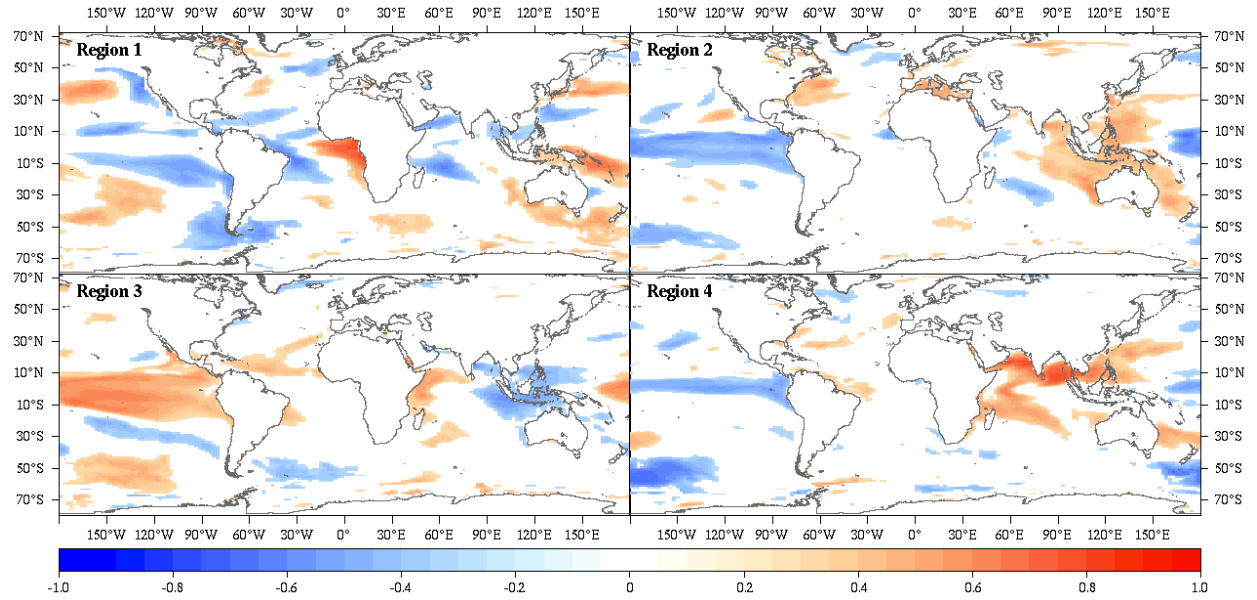
Fig. 10 Correlation patterns of CHIRPS precipitation with global ERSST for the mean timeseries of the four regions of interannual variability of JAS precipitation at the period (1981-2014). All correlations are significant at 90% confidence level.

565

Fig. 11 Correlation patterns of CCSM4 precipitation with the model SST for the mean timeseries

of the four regions of interannual variability of JAS precipitation at the period (1960-1990). All

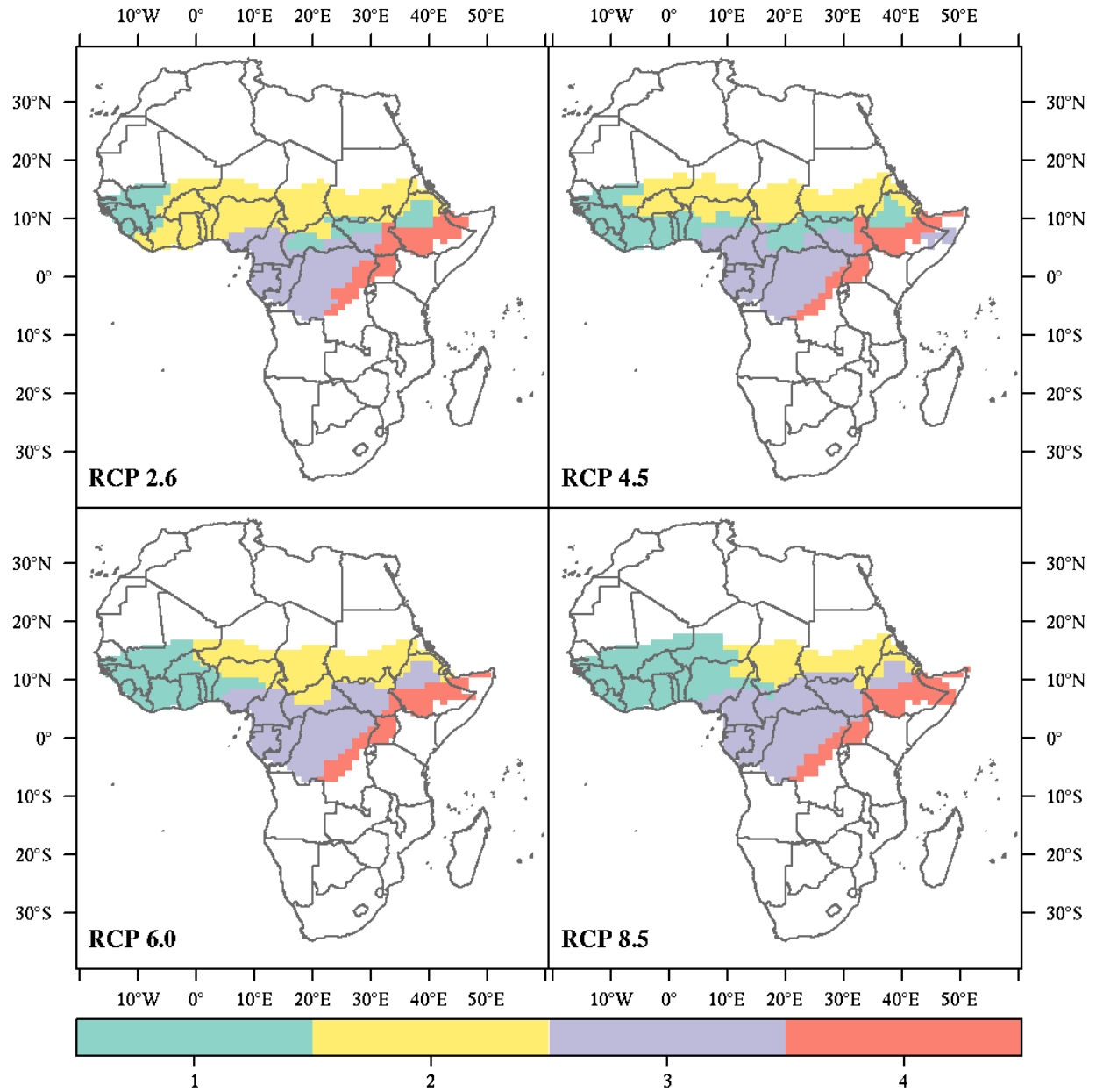correlations are significant at 90% confidence level.

569

Fig. 12 Changes in the regions of JAS precipitation over Africa using regional-linkage method for different CCSM4 climate projections at the entire simulation period (2006-2100).
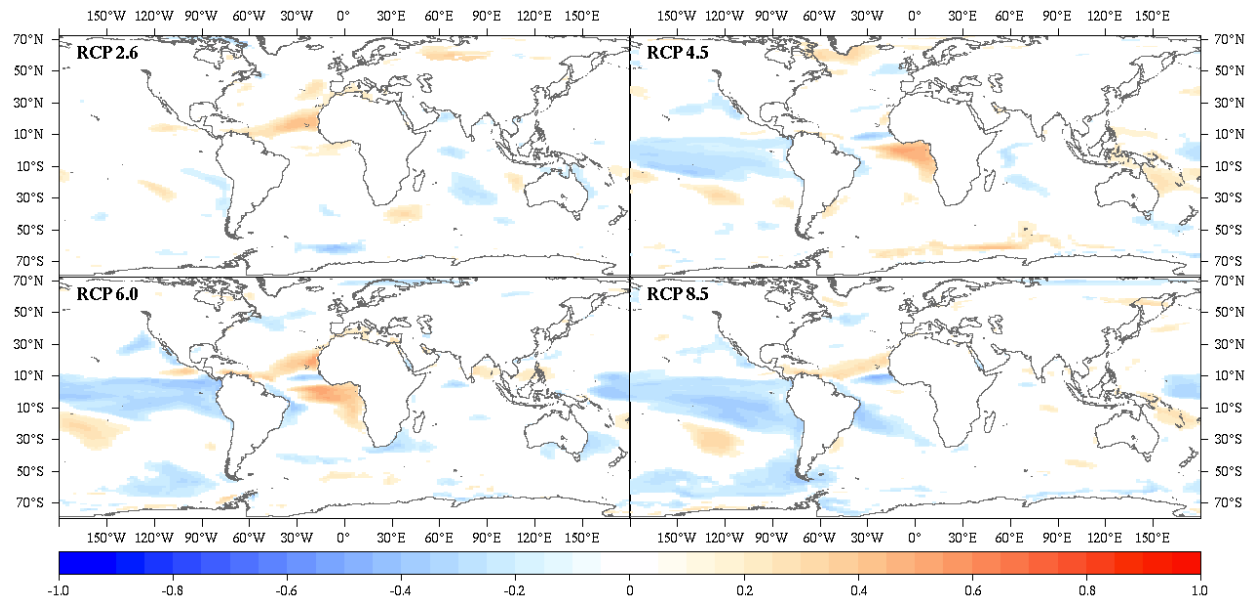
Fig. 13 Correlation patterns of CCSM4 JAS precipitation from the four RCPs with the corresponding SST for region 1 (Western Sahel in **Fig. 12**) at the period (2006-2100). All correlations are significant at 90% confidence level.
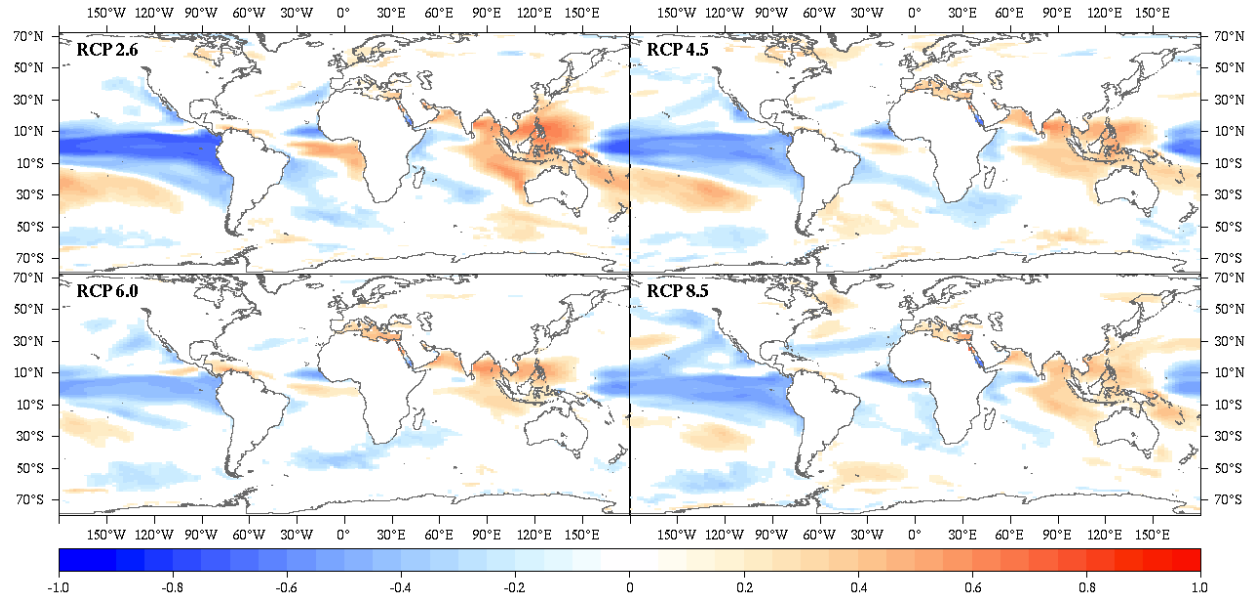
Fig. 14 Correlation patterns of CCSM4 JAS precipitation from the four RCPs with the corresponding SST for region 2 (Eastern Sahel in **Fig. 12**) at the period (2006-2100). All correlations are significant at 90% confidence level.